GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:31 ; Search time 91.75 Seconds
(without alignments)
22.175 Million cell updates/sec

Title: US-09-331-631A-1_COPY_186_248
Perfect score: 355
Sequence: 1 KRDPQQREYEDCRRRCEQQE.....MMNPQRGGSGRYEEGEEEQS 63

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	114	32.1	588	1	VCLB_GOSHI	P09801 gossypium h
2	101	28.5	605	1	VCLA_GOSHI	P09799 gossypium h
3	84	23.7	966	1	SSN6_YEAST	P14922 saccharomyc
4	79	22.3	919	1	ANDR_HUMAN	P10275 homo sapien
5	75	21.1	1090	1	NIT4_NEUCR	P28349 neurospora
6	74	20.8	1898	1	TRHY_HUMAN	Q07283 homo sapien
7	74	20.8	2442	1	CBP_HUMAN	Q92793 homo sapien
8	73.5	20.7	185	1	T2_MOUSE	Q06666 mus musculu
9	73	20.6	573	1	GLB1_MAIZE	P15590 zea mays (m
10	73	20.6	911	1	ANDR_PANTR	O97775 pan troglod
11	70.5	19.9	1344	1	IF3A_MOUSE	P23116 mus musculu
12	69	19.4	239	1	WT1_SMIMA	P49953 sminthopsis
13	69	19.4	288	1	WT1_ALLMI	P50902 alligator m
14	69	19.4	448	1	WT1_RAT	P49952 rattus norv
15	69	19.4	449	1	WT1_HUMAN	P19544 homo sapien
16	69	19.4	449	1	WT1_MOUSE	P22561 mus musculu
17	69	19.4	449	1	WT1_PIG	O62651 sus scrofa
18	69	19.4	895	1	ANDR_PAPHA	O97960 papio hamad
19	69	19.4	1023	1	GLT_DROME	P33438 drosophila
20	68.5	19.3	758	1	YM38_YEAST	Q03825 saccharomyc
21	68.5	19.3	1382	1	IF3A_HUMAN	Q14152 homo sapien
22	68	19.2	1154	1	WC1_NEUCR	Q01371 neurospora
23	68	19.2	1265	1	CYA5_CANFA	P30803 canis famil
24	67.5	19.0	164	1	2SS3_ARATH	P15459 arabidopsis
25	66.5	18.7	170	1	2SS2_ARATH	P15458 arabidopsis
26	66.5	18.7	372	1	HXA2_MOUSE	P31245 mus musculu
27	66.5	18.7	372	1	HXA2_RAT	P31246 rattus norv
28	66	18.6	423	1	TF3A_HUMAN	Q92664 homo sapien
29	66	18.6	529	1	GAG_MLVDU	P23090 duplan muri
30	66	18.6	536	1	GAG_MLVDE	P29168 murine leuk
31	66	18.6	538	1	GAG_MLVFP	P26805 friend muri
32	66	18.6	539	1	GAG_MLVF5	P26807 friend muri
33	66	18.6	540	1	GAG_MLVHO	P21435 homulv muri
34	66	18.6	614	1	RU17_HUMAN	P08621 homo sapien
35	66	18.6	1407	1	TRHY_RABIT	P37709 oryctolagus
36	66	18.6	1596	1	MAM_DROME	P21519 drosophila
37	65.5	18.5	47	1	AGRP_LUFCY	P56568 luffa cylin
38	65.5	18.5	255	1	LP61_EIMTE	P15714 eimeria ten
39	65.5	18.5	450	1	SRF_DROME	Q24535 drosophila
40	65.5	18.5	568	1	INVO_RAT	P48998 rattus norv
41	65.5	18.5	2124	1	Y192_HUMAN	Q93074 homo sapien
42	65	18.3	280	1	YTZ4_AGRVI	P70796 agrobacteri
43	65	18.3	339	1	TF2D_HUMAN	P20226 homo sapien
44	65	18.3	384	1	INVO_GALCR	P24710 galago cras
45	65	18.3	905	1	SNF5_YEAST	P18480 saccharomyc

ALIGNMENTS

RESULT 1
VCLB_GOSHI
ID VCLB_GOSHI STANDARD; PRT; 588 AA.
AC P09801;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XVIII. cDNA and amino acid sequences of the members of
RT the storage protein families.";
RL Plant Mol. Biol. 7:475-489(1986).
CC -!- FUNCTION: SEED STORAGE PROTEIN.
CC -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC ---

CC ---
DR EMBL; M16891; AAA33071.1; -.
DR PIR; A30838; FWCNAB.
DR HSSP; P50477; 1CAX.
DR INTERPRO; IPR001113; -.
DR PFAM; PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 588 VICILIN C72.
SQ SEQUENCE 588 AA; 69729 MW; 63E699B29AB8ADEB CRC64;

Query Match 32.1%; Score 114; DB 1; Length 588;
Best Local Similarity 33.7%; Pred. No. 9.6e-05;
Matches 33; Conservative 9; Mismatches 18; Indels 38; Gaps 5;

```
Qy       3 DPQQREYEDCRRRCEQQEPRQQHQCQLRC-----REQQ---------RQHGRGGDMM--- 45
           |||:| ||:|:: | ||| ||: ||| ||     :|||          :||    :
Db      82 DPQRR-YEECQQECRQQEERQRPQCQQRCLKRFEQEQQQSQRQFQECQQHCHQQEQRPER 140

Qy      46 --------------NPQR------GGSGRYEEGEEEQS 63
                         || |            ||||:|||
Db     141 KQQCVRECRERYQENPWRREREEEAEEEETEEGEQEQS 178
```

```
RESULT   2
VCLA_GOSHI
ID   VCLA_GOSHI     STANDARD;      PRT;   605 AA.
AC   P09799;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XIX. Sequences and genomic organization of the alpha
RT   globulin (vicilin) genes of cottonseed.";
RL   Plant Mol. Biol. 9:533-546(1987).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19378; AAA33069.1; -.
DR   PIR; S06398; S06398.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL        1     23
FT   CHAIN        24    605       VICILIN GC72-A.
SQ   SEQUENCE   605 AA;  71049 MW;  C9DB9371C976953B CRC64;


  Query Match             28.5%;  Score 101;  DB 1;  Length 605;
  Best Local Similarity   27.7%;  Pred. No. 0.002;
  Matches   26;  Conservative   15;  Mismatches   19;  Indels   34;  Gaps    4;

Qy       3 DPQQREYEDCRRRCEQQEPRQQHQCQLRCREQ-QRQHGRGGDMM---------------- 45
           |||:| |:|||: |:|:| | :  |:  |||| ::|  :  |
Db      79 DPQRR-YQDCRQHCQQEERRLRPHCEQSCREQYEKQQQQQQPDKQFKECQQRCQWQEQRPE 137

Qy      46 ---------------NPQRG-GSGRYEEGEEEQS 63
                          :| :|    :: | |||:|
Db     138 RKQQCVKECREQYQEDPWKGERENKWREEEEEES 171


RESULT   3
SSN6_YEAST
ID   SSN6_YEAST     STANDARD;      PRT;   966 AA.
AC   P14922;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   GLUCOSE REPRESSION MEDIATOR PROTEIN.
GN   SSN6 OR CYC8 OR YBR112C OR YBR0908.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89211964; PubMed=2854095;
RA   Trumbly R.J.;
RT   "Cloning and characterization of the CYC8 gene mediating glucose
RT   repression in yeast.";
RL   Gene 73:97-111(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88065502; PubMed=3316983;
RA   Schultz J., Carlson M.;
RT   "Molecular analysis of SSN6, a gene functionally related to the SNF1
RT   protein kinase of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:3637-3645(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [4]
RP   TPR REPEATS.
RX   MEDLINE=90124639; PubMed=2404612;
RA   Sikorski R.S., Boguski M.S., Goebl M., Hieter P.A.;
RT   "A repeating amino acid motif in CDC23 defines a family of proteins
RT   and a new relationship among genes required for mitosis and RNA
RT   synthesis.";
RL   Cell 60:307-317(1990).
CC   -!- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND
CC       IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC       PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC       DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 10 TPR DOMAINS.
CC   -!- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23440; AAA34545.1; -.
DR   EMBL; M17826; AAA35103.1; -.
DR   EMBL; X66247; CAA46973.1; -.
DR   EMBL; X78993; CAA55615.1; -.
DR   EMBL; Z35981; CAA85069.1; -.
DR   PIR; S25365; S25365.
DR   SGD; S0000316; CYC8.
DR   INTERPRO; IPR001440; -.
DR   PFAM; PF00515; TPR; 7.
KW   Transcription regulation; Repressor; Repeat; TPR domain;
KW   Nuclear protein.
FT   DOMAIN       15     30       POLY-GLN.
FT   REPEAT       46     79       TPR 1.
FT   REPEAT       80    113       TPR 2.
FT   REPEAT      114    147       TPR 3.
FT   REPEAT      150    183       TPR 4.
FT   REPEAT      187    220       TPR 5.
FT   REPEAT      224    257       TPR 6.
FT   REPEAT      258    291       TPR 7.
FT   REPEAT      296    329       TPR 8.
FT   REPEAT      330    363       TPR 9.
FT   REPEAT      364    398       TPR 10.
FT   DOMAIN      493    556       30 X 2 AA TANDEM REPEATS OF Q-A.
FT   DOMAIN      557    587       POLY-GLN.
FT   CONFLICT    547    547       K -> Q (IN REF. 3).
SQ   SEQUENCE   966 AA;  107202 MW;  84B509CF3208C5C0 CRC64;


  Query Match             23.7%;  Score 84;  DB 1;  Length 966;
```

```
  Best Local Similarity   34.0%;  Pred. No. 0.16;
  Matches   17;  Conservative   15;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       1 KRDPQQREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMMNPQRG 50
           ::  ||:: :  ::: :||: :|| | |   |:| :| |    |:|||:|
Db     562 QQQQQQQQQQQQQQQQQQQQQQQQQQLQPLPRQQLQQKGVSVQMLNPQQG 611


RESULT   4
ANDR_HUMAN
ID   ANDR_HUMAN     STANDARD;      PRT;   919 AA.
AC   P10275;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
GN   AR OR NR3C4 OR DHTR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89112208; PubMed=3216866;
RA   Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,
RA   French F.S., Wilson E.M.;
RT   "The human androgen receptor: complementary deoxyribonucleic acid
RT   cloning, sequence analysis and gene expression in prostate.";
RL   Mol. Endocrinol. 2:1265-1275(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90083302; PubMed=2594783;
RA   Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
RA   Wilson E.M., French F.S.;
RT   "Sequence of the intron/exon junctions of the coding region of the
RT   human androgen receptor gene and identification of a point mutation
RT   in a family with complete androgen insensitivity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258935; PubMed=2342476;
RA   Govindan M.V.;
RT   "Specific region in hormone binding domain is essential for hormone
RT   binding and trans-activation by human androgen receptor.";
RL   Mol. Endocrinol. 4:417-427(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATE;
RX   MEDLINE=89017168; PubMed=3174628;
RA   Chang C., Kokontis J., Liao S.;
RT   "Structural analysis of complementary DNA and amino acid sequences of
RT   human and rat androgen receptors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATE;
RX   MEDLINE=89098909; PubMed=2911578;
RA   Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT   "Characterization and expression of a cDNA encoding the human androgen
RT   receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATE;
RX   MEDLINE=91155943; PubMed=2293020;
RA   Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RA   McPhaul M.J.;
RT   "Definition of the human androgen receptor gene structure permits the
RT   identification of mutations that cause androgen resistance: premature
RT   termination of the receptor protein at amino acid residue 588 causes
RT   complete androgen resistance.";
RL   Mol. Endocrinol. 4:1105-1116(1990).
RN   [7]
RP   SEQUENCE OF 189-919 FROM N.A.
RX   MEDLINE=88178111; PubMed=3353726;
RA   Chang C., Kokontis J., Liao S.;
RT   "Molecular cloning of human and rat complementary DNA encoding
RT   androgen receptors.";
RL   Science 240:324-326(1988).
RN   [8]
RP   SEQUENCE OF 468-919 FROM N.A.
RX   MEDLINE=88240407; PubMed=3377788;
RA   Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RA   Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RA   Mulder E., Brinkmann A.O.;
RT   "Cloning, structure and expression of a cDNA encoding the human
RT   androgen receptor.";
RL   Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN   [9]
RP   POLYMORPHISM OF POLY-GLN REGION.
RX   MEDLINE=92220629; PubMed=1561105;
RA   Sleddens H.F., Oostra B.A., Brinkmann A.O., Trapman J.;
RT   "Trinucleotide repeat polymorphism in the androgen receptor gene
RT   (AR).";
RL   Nucleic Acids Res. 20:1427-1427(1992).
RN   [10]
RP   POLYMORPHISM OF POLY-GLY REGION.
RC   TISSUE=BLOOD;
RA   Lu J., Danielsen M.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   VARIANTS SBMA IN POLY-GLN REGION.
RX   MEDLINE=91287825; PubMed=2062380;
RA   la Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fischbeck K.H.;
RT   "Androgen receptor gene mutations in X-linked spinal and bulbar
RT   muscular atrophy.";
RL   Nature 352:77-79(1991).
RN   [12]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=95023089; PubMed=7937057;
RA   Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;
RT   "The androgen receptor gene mutations database.";
RL   Nucleic Acids Res. 22:3560-3562(1994).
RN   [13]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97169385; PubMed=9016528;
RA   Gottlieb B., Trifiro M., Lumbroso R., Vasiliou D.M., Pinsky L.;
RT   "The androgen receptor gene mutations database.";
RL   Nucleic Acids Res. 25:158-162(1997).
RN   [14]
RP   VARIANT LNCAP ALA-877.
RX   MEDLINE=91083633; PubMed=2260966;
RA   Veldscholte J., Ris-Stalpers C., Kuiper G.G., Jenster G.,
RA   Berrevoets C., Claassen E., van Rooij H.C.J., Trapman J.,
RA   Brinkmann A.O., Mulder E.;
RT   "A mutation in the ligand binding domain of the androgen receptor of
RT   human LNCaP cells affects steroid binding characteristics and
RT   response to anti-androgens.";
RL   Biochem. Biophys. Res. Commun. 173:534-540(1990).
RN   [15]
RP   VARIANT CAIS MET-866.
RX   MEDLINE=91186983; PubMed=2082179;
RA   Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
RA   Corfen J.L.;
RT   "Functional characterization of naturally occurring mutant androgen
RT   receptors from subjects with complete androgen insensitivity.";
RL   Mol. Endocrinol. 4:1759-1772(1990).
RN   [16]
RP   VARIANT CYS-774.
RX   MEDLINE=91310758; PubMed=1856263;
RA   Marcelli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
RA   McPhaul M.J.;
RT   "Androgen resistance associated with a mutation of the androgen
RT   receptor at amino acid 772 (Arg-->Cys) results from a combination of
RT   decreased messenger ribonucleic acid levels and impairment of
RT   receptor function.";
RL   J. Clin. Endocrinol. Metab. 73:318-325(1991).
```

```
RN   [17]
RP   VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.
RX   MEDLINE=92131007; PubMed=1775137;
RA   Ris-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G.,
RA   Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Liao S.;
RT   "Substitution of aspartic acid-686 by histidine or asparagine in the
RT   human androgen receptor leads to a functionally inactive protein with
RT   altered hormone-binding characteristics.";
RL   Mol. Endocrinol. 5:1562-1569(1991).
RN   [18]
RP   VARIANTS CAIS AND PAIS.
RX   MEDLINE=93338440; PubMed=1307250;
RA   Batch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,
RA   Hughes I.A., Patterson M.N.;
RT   "Androgen receptor gene mutations identified by SSCP in fourteen
RT   subjects with androgen insensitivity syndrome.";
RL   Hum. Mol. Genet. 1:497-503(1992).
RN   [19]
RP   VARIANT CAIS VAL-787.
RX   MEDLINE=92235226; PubMed=1569163;
RA   Nakao R., Haji M., Yanase T., Ogo A., Takayanagi R., Katsube T.,
RA   Fukumaki Y., Nawata H.;
RT   "A single amino acid substitution (Met-786-->Val) in the steroid-
RT   binding domain of human androgen receptor leads to complete androgen
RT   insensitivity syndrome.";
RL   J. Clin. Endocrinol. Metab. 74:1152-1157(1992).
RN   [20]
RP   VARIANT LNCAP ALA-877.
RX   MEDLINE=92222955; PubMed=1562539;
RA   Veldscholte J., Berrevoets C.A., Ris-Stalpers C., Kuiper G.G.,
RA   Jenster G., Trapman J., Brinkmann A.O., Mulder E.;
RT   "The androgen receptor in LNCaP cells contains a mutation in the
RT   ligand binding domain which affects steroid binding characteristics
RT   and response to antiandrogens.";
RL   J. Steroid Biochem. Mol. Biol. 41:665-669(1992).
RN   [21]
RP   VARIANT MET-730.
RX   MEDLINE=92335289; PubMed=1631125;
RA   Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,
RA   Isaacs W.B., Brown T.R., Barrack E.R.;
RT   "Androgen receptor gene mutations in human prostate cancer.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).
RN   [22]
RP   VARIANT CAIS VAL-754.
RX   MEDLINE=93372806; PubMed=8103398;
RA   Lobaccaro J.-M., Lumbroso S., Ktari R., Dumas R., Sultan C.;
RT   "An exonic point mutation creates a MaeIII site in the androgen
RT   receptor gene of a family with complete androgen insensitivity
RT   syndrome.";
RL   Hum. Mol. Genet. 2:1041-1043(1993).
RN   [23]
RP   VARIANT CAIS ARG-807.
RX   MEDLINE=94108430; PubMed=8281140;
RA   Adeyemo O., Kallio P.J., Palvimo J.J., Kontula K., Jaenne O.A.;
RT   "A single-base substitution in exon 6 of the androgen receptor gene
RT   causing complete androgen insensitivity: the mutated receptor fails
RT   to transactivate but binds to DNA in vitro.";
RL   Hum. Mol. Genet. 2:1809-1812(1993).
RN   [24]
RP   VARIANT PAIS VAL-743.
RX   MEDLINE=93315568; PubMed=8325932;
RA   Nakao R., Yanase T., Sakai Y., Haji M., Nawata H.;
RT   "A single amino acid substitution (Gly743 --> Val) in the steroid-
RT   binding domain of the human androgen receptor leads to Reifenstein

  Query Match             22.3%;  Score 79;  DB 1;  Length 919;
  Best Local Similarity   32.8%;  Pred. No. 0.51;
  Matches   19;  Conservative   15;  Mismatches   22;  Indels    2;  Gaps    1;

Qy       5 QQREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMMNPQRGGSGRYEEGEEEQ 62
           ||:: :  ::: :||: :|| | :   |:||:| |  |     :|| :|    ||:|
Db      58 QQQQQQ--QQQQQQQQQQQQQQQETSPRQQQQQQGEDGSPQAHRRGPTGYLVLDEEQQ 113
```

```
RESULT   5
NIT4_NEUCR
ID   NIT4_NEUCR     STANDARD;      PRT;  1090 AA.
AC   P28349;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIT-4.
GN   NIT-4.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92017855; PubMed=1840634;
RA   Yuan G.-F., Fu Y.-H., Marzluf G.A.;
RT   "nit-4, a pathway-specific regulatory gene of Neurospora crassa,
RT   encodes a protein with a putative binuclear zinc DNA-binding
RT   domain.";
RL   Mol. Cell. Biol. 11:5735-5745(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92149315; PubMed=1531376;
RA   Yuan G.-F., Marzluf G.A.;
RT   "Molecular characterization of mutations of nit-4, the
RT   pathway-specific regulatory gene which controls nitrate assimilation
RT   in Neurospora crassa.";
RL   Mol. Microbiol. 6:67-73(1992).
CC   -!- FUNCTION: PATHWAY-SPECIFIC REGULATORY GENE OF NITRATE
CC       ASSIMILATION; IT ACTIVATES THE TRANSCRIPTION OF THE GENES FOR
CC       NITRATE AND NITRITE REDUCTASES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING
CC       GENE EXPRESSION.
CC   -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC       CLUSTER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80368; AAA33602.1; -.
DR   PIR; A41696; A41696.
DR   PIR; S20033; S20033.
DR   HSSP; P07272; 1PYI.
DR   INTERPRO; IPR001138; -.
DR   PFAM; PF00172; Zn_clus; 1.
DR   PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR   PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Zinc; Metal-binding; Nitrate assimilation.
FT   DNA_BIND     53     81       ZN(2)-CYS(6), FUNGAL-TYPE.
FT   DOMAIN      121    139       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      213    229       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      429    450       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      672    754       PRO-RICH.
FT   DOMAIN      755    859       GLN-RICH.
FT   DOMAIN      992   1024       POLY-GLN.
FT   CONFLICT     98     98       K -> KP (IN REF. 1).
FT   CONFLICT    467    467       L -> S (IN REF. 1).
SQ   SEQUENCE   1090 AA;  120244 MW;  881D89172EDD6114 CRC64;

  Query Match             21.1%;  Score 75;  DB 1;  Length 1090;
  Best Local Similarity   32.7%;  Pred. No. 1.5;
  Matches   17;  Conservative   14;  Mismatches   17;  Indels    4;  Gaps    1;

Qy       6 QREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMM----NPQRGGSG 53
           ||: :  |:: :||: :|| | | : ::||:|  :  :|       :||| |
```

```
Db    990 QRQQQQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQEANMFAYIHHGAERGGGG 1041


RESULT   6
TRHY_HUMAN
ID   TRHY_HUMAN     STANDARD;      PRT;  1898 AA.
AC   Q07283;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRICHOHYALIN.
GN   THH OR TRHY OR THL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93280194; PubMed=7685034;
RA   Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA   Steinert P.M.;
RT   "The structure of human trichohyalin. Potential multiple roles as a
RT   functional EF-hand-like calcium-binding protein, a cornified cell
RT   envelope precursor, and an intermediate filament-associated (cross-
RT   linking) protein.";
RL   J. Biol. Chem. 268:12164-12176(1993).
RN   [2]
RP   SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=93315897; PubMed=7686953;
RA   O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RT   "Trichohyalin: a structural protein of hair, tongue, nail, and
RT   epidermis.";
RL   J. Invest. Dermatol. 101:65S-71S(1993).
CC   -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC       IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC       INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC       LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC       ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC       WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC       ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC       ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC       DIFFERENTIATION.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC       THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC       THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC       THE EPIDERMIS.
CC   -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC       CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC       ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC       OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC       ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC       THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC       DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC       THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC       DIFFERENT SPECIES.
CC   -!- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC       PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09190; AAA65582.1; -.
DR   PIR; A45973; A45973.
DR   HSSP; P02633; 1BOC.
DR   MIM; 190370; -.
DR   INTERPRO; IPR001751; -.
DR   INTERPRO; IPR002048; -.
DR   PFAM; PF01023; S_100; 1.
DR   PFAM; PF00036; efhand; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Repeat; Calcium-binding.
FT   DOMAIN        1     91       S-100 LIKE.
FT   CA_BIND      22     33       SITE I (LOW AFFINITY) (POTENTIAL).
FT   CA_BIND      62     73       SITE II (HIGH AFFINITY) (POTENTIAL).
FT   DOMAIN      314    390       6 X 13 AA TANDEM REPEATS OF
FT                                R-R-E-Q-E-E-E-R-R-E-Q-Q-L.
FT   REPEAT      314    326       1-1 (APPROXIMATE).
FT   REPEAT      327    339       1-2 (APPROXIMATE).
FT   REPEAT      340    351       1-3 (APPROXIMATE).
FT   REPEAT      352    364       1-4.
FT   REPEAT      365    377       1-5.
FT   REPEAT      378    390       1-6.
FT   DOMAIN      391    444       9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT   REPEAT      391    396       2-1.
FT   REPEAT      397    402       2-2.
FT   REPEAT      403    408       2-3.
FT   REPEAT      409    414       2-4.
FT   REPEAT      415    420       2-5.
FT   REPEAT      421    426       2-6.
FT   REPEAT      427    432       2-7.
FT   REPEAT      433    438       2-8.
FT   REPEAT      439    444       2-9.
FT   DOMAIN      444    702       9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT   DOMAIN      923   1162       8 X 30 AA TANDEM REPEATS.
FT   REPEAT      923    952       4-1.
FT   REPEAT      953    982       4-2.
FT   REPEAT      983   1012       4-3.
FT   REPEAT     1013   1042       4-4.
FT   REPEAT     1043   1072       4-5.
FT   REPEAT     1073   1102       4-6.
FT   REPEAT     1103   1132       4-7.
FT   REPEAT     1133   1162       4-8.
FT   DOMAIN     1250   1849       23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT   CONFLICT   1752   1752       F -> L (IN REF. 2).
FT   CONFLICT   1794   1801       QERDRQYR -> RSETGSTG (IN REF. 2).
FT   CONFLICT   1857   1857       Q -> K (IN REF. 2).
FT   CONFLICT   1880   1880       V -> G (IN REF. 2).
SQ   SEQUENCE   1898 AA;  247219 MW;  A74B5947FB62E31D CRC64;


  Query Match             20.8%;  Score 74;  DB 1;  Length 1898;
  Best Local Similarity   36.4%;  Pred. No. 3.1;
  Matches   24;  Conservative   10;  Mismatches   28;  Indels    4;  Gaps    2;

Qy       1 KRDPQQREYEDCRRRCEQQEPR-QQHQCQLRCREQQRQHGRGGDMMNPQRGGSGRYEEGE 59
           :|| : || |  |:  |:|: | |:   : |  || ||      :   || |  |:|| :
Db    1723 ERDRKFREEEQLRQGREEQQLRSQESDRKFREEEQLRQEREEQQLRPQQRDGKYRWEEEQ 1782

Qy      60 ---EEQ 62
              |||
Db    1783 LQLEEQ 1788


RESULT   7
CBP_HUMAN
ID   CBP_HUMAN      STANDARD;      PRT;  2442 AA.
AC   Q92793; Q16376; O00147;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CREB-BINDING PROTEIN.
GN   CREBBP OR CBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97385172; PubMed=9238046;
RA   Sobulo O.M., Borrow J., Tomek R., Reshimi S., Harden A.,
RA   Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA   Zeleznik-Le N.J.;
RT   "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT   acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97321049; PubMed=9177780;
RA   Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA   Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA   Doggett N.A., Peters D.J.M., Breuning M.H.;
RT   "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT   of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT   16p13.3.";
RL   Genomics 42:96-144(1997).
RN   [3]
RP   SEQUENCE OF 1-405 FROM N.A.
RX   MEDLINE=96376968; PubMed=8782817;
RA   Borrow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
RA   Chaganti R.S.K., Civin C.I., Disteche C., Dube I., Frischauf A.M.,
RA   Horsman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT   "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT   a putative acetyltransferase to the CREB-binding protein.";
RL   Nat. Genet. 14:33-41(1996).
CC   -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC       PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC       THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC       CAMP-RESPONSIVE GENES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC       T(8;16)(P11;P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC       INVOLVING CBP AND MLL.
CC   -!- DISEASE: DEFECTS IN CBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC       SYNDROME (RTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC       ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC       AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC   -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47741; AAC51770.1; -.
DR   EMBL; U85962; AAC51331.1; -.
DR   EMBL; U89354; AAC51339.1; -.
DR   EMBL; U89355; AAC51340.1; -.
DR   MIM; 600140; -.
DR   INTERPRO; IPR000197; -.
DR   INTERPRO; IPR000433; -.
DR   INTERPRO; IPR001487; -.
DR   PFAM; PF00569; ZZ; 1.
DR   PFAM; PF00439; bromodomain; 1.
DR   PFAM; PF02135; zf-TAZ; 2.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
KW   Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW   Chromosomal translocation.
FT   DOMAIN      363    430       CYS/HIS-RICH.
FT   DOMAIN      452    683       CREB-BINDING.
FT   DOMAIN     1103   1175       BROMODOMAIN.
FT   DOMAIN     1061   1064       POLY-GLU.
FT   DOMAIN     1199   1487       CYS/HIS-RICH.
FT   DOMAIN     1555   1562       POLY-GLU.
FT   DOMAIN     1675   1849       CYS/HIS-RICH.
FT   DOMAIN     1943   1948       POLY-PRO.
FT   DOMAIN    -1967   1970       POLY-GLN.
FT   DOMAIN     2081   2085       POLY-GLN.
FT   DOMAIN     2199   2216       POLY-GLN.
FT   DOMAIN     2245   2248       POLY-GLN.
FT   DOMAIN     2297   2300       POLY-GLN.
FT   CONFLICT   1511   1513       FAE -> NSG (IN REF. 2).
FT   CONFLICT   1724   1725       ED -> VV (IN REF. 2).
FT   CONFLICT   1770   1770       V -> L (IN REF. 2).
FT   CONFLICT   1789   1789       N -> F (IN REF. 2).
FT   CONFLICT   1812   1812       T -> P (IN REF. 2).
SQ   SEQUENCE   2442 AA;  265336 MW;  42D084619475F3D2 CRC64;


  Query Match             20.8%;  Score 74;  DB 1;  Length 2442;
  Best Local Similarity   34.8%;  Pred. No. 4;
  Matches   23;  Conservative    9;  Mismatches   24;  Indels   10;  Gaps    3;

Qy       3 DPQQREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDM------MNPQRGGSGRYE 56
           :|| ||    ||:  ||: :|| | | : ::||   |  | |        || | | |
Db    2186 NPQYREM--LRRQLLQQQQQQQQQQQQQQQQQQGSAGMAGGMAGHGQFQQPQ--GPGGYP 2241

Qy      57 EGEEEQ 62
             ::|
Db    2242 PAMQQQ 2247


RESULT   8
T2_MOUSE
ID   T2_MOUSE       STANDARD;      PRT;   185 AA.
AC   Q06666;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   OCTAPEPTIDE-REPEAT PROTEIN T2.
GN   SRST OR T2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=MACROPHAGE;
RX   MEDLINE=93092084; PubMed=1458435;
RA   di Carlo M., Montana G., Romancino D.P., Monteleone D.;
RT   "A mouse repeat sequence conserved in eukaryotic genomes.";
RL   J. Submicrosc. Cytol. Pathol. 24:467-472(1992).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67863; CAA48048.1; -.
DR   MGD; MGI:107677; Srst.
KW   Repeat.
FT   DOMAIN        2    159       20 X 8 AA APPROXIMATE TANDEM REPEATS.
SQ   SEQUENCE   185 AA;  22805 MW;  8995BEC4EC383971 CRC64;


  Query Match             20.7%;  Score 73.5;  DB 1;  Length 185;
  Best Local Similarity   29.0%;  Pred. No. 0.42;
  Matches   18;  Conservative   14;  Mismatches   25;  Indels    5;  Gaps    2;

Qy       1 KRDPQQREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMMNPQRGGSGRYEEGEE 60
             :|: :::   :  |  |:|  ||:   : | || :|| ||  :    |||     : ||
Db      96 EREAERQRGREAEREAERQRERQRE--RQRGREAERQRGREAER---QRGREAEKQRGER 150

Qy      61 EQ 62
           ::
Db     151 QR 152
```

```
RESULT  9
GLB1_MAIZE
ID   GLB1_MAIZE     STANDARD;      PRT;   573 AA.
AC   P15590;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-LIKE).
GN   GLB1.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. INBRED LINE VA26;
RA   Belanger F.C., Kriz A.L.;
RT   "Molecular characterization of the major maize embryo globulin encoded
RT   by the Glb1 gene.";
RL   Plant Physiol. 91:636-643(1989).
RN   [2]
RP   SEQUENCE OF 87-100.
RX   MEDLINE=89374022; PubMed=2775172;
RA   Kriz A.L.;
RT   "Characterization of embryo globulins encoded by the maize Glb
RT   genes.";
RL   Biochem. Genet. 27:239-251(1989).
CC   -!- PTM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
CC       MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
CC   -!- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLB1 ALLELES HAVE
CC       THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
CC       PROTEINS, RESPECTIVELY.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M24845; AAA33467.1; -.
DR   HSSP; P50477; 1CAW.
DR   MAIZEDB; 30181; -.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL        1     18       OR 21 (POTENTIAL).
FT   PROPEP       19     86
FT   CHAIN        87    573       GLOBULIN-1 S ALLELE.
FT   CARBOHYD    349    349       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   573 AA;  65029 MW;  525ED1D00A062976 CRC64;


  Query Match             20.6%;  Score 73;  DB 1;  Length 573;
  Best Local Similarity   32.1%;  Pred. No. 1.3;
  Matches   17;  Conservative   10;  Mismatches   24;  Indels    2;  Gaps    1;

Qy      12 CRRRCEQQEPRQQHQCQLRCREQQRQ--HGRGGDMMNPQRGGSGRYEEGEEEQ 62
           | ||||  :   |: :|  :|||::|:    |     : : |     :| | ||
Db      40 CVRRCEDRPWHQRPRCLEQCREEEREKRQERSRHEADDRSGEGSSEDERERQ 92


RESULT  10
ANDR_PANTR
ID   ANDR_PANTR     STANDARD;      PRT;   911 AA.
AC   O97775;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
GN   AR OR NR3C4.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98404153; PubMed=9732460;
RA   Choong C.S., Kemppainen J.A., Wilson E.M.;
RT   "Evolution of the primate androgen receptor: a structural basis for
RT   disease.";
RL   J. Mol. Evol. 47:334-342(1998).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U94177; AAC73048.1; -.
DR   HSSP; P06536; 1RGD.
DR   INTERPRO; IPR000536; -.
DR   INTERPRO; IPR001103; -.
DR   INTERPRO; IPR001628; -.
DR   PFAM; PF00104; hormone_rec; 1.
DR   PFAM; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   PRINTS; PR00521; ANDROGENR.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    549       MODULATING (BY SIMILARITY).
FT   DNA_BIND    551    616       C4-TYPE ZINC FINGERS (TWO).
FT   ZN_FING     551    571       C4-TYPE.
FT   ZN_FING     587    611       C4-TYPE.
FT   DOMAIN      682    911       LIGAND-BINDING.
FT   DOMAIN       57     78       POLY-GLN.
FT   DOMAIN       84     88       POLY-GLN.
FT   DOMAIN      192    196       POLY-GLN.
FT   DOMAIN      371    380       POLY-PRO.
FT   DOMAIN      395    401       POLY-ALA.
FT   DOMAIN      448    464       POLY-GLY.
SQ   SEQUENCE   911 AA;  98402 MW;  601B9BD4E697DAA4 CRC64;


  Query Match             20.6%;  Score 73;  DB 1;  Length 911;
  Best Local Similarity   32.8%;  Pred. No. 2;
  Matches   19;  Conservative   13;  Mismatches   24;  Indels    2;  Gaps    1;

Qy       5 QQREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMMNPQRGGSGRYEEGEEEQ 62
             ||:: :  ::: :||: :|| | |     ||:| |  |     :|| :|     ||:|
Db      57 QQQQQQ--QQQQQQQQQQQQQQQETSPRQQQQQGEDGSPQAHRRGPTGYLVLDEEQQ 112


RESULT  11
IF3A_MOUSE
ID   IF3A_MOUSE     STANDARD;      PRT;  1344 AA.
AC   P23116; Q60697; Q62162;
DT   01-NOV-1991 (Rel. 20, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT 10 (EIF-3 THETA)
DE   (EIF3 P167) (EIF3 P180) (EIF3 P185) (P162 PROTEIN) (CENTROSOMIN).
GN   EIF3S10 OR EIF3 OR CSMA.
```

```
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LYMPHOMA;
RA   Fisher R., Fillmore H., Reynolds A.B.;
RT   "Molecular cloning and characterization of the 162 kDa component
RT   of a multi-protein complex phosphorylated by Src.";
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 514-961 FROM N.A. (CENTROSOMIN B).
RX   MEDLINE=98039715; PubMed=9372446;
RA   Petzelt C., Joswig G., Mincheva A., Lichter P., Stammer H., Werner D.;
RT   "The centrosomal protein centrosomin A and the nuclear protein
RT   centrosomin B derive from one gene by post-transcriptional processes
RT   involving RNA editing.";
RL   J. Cell Sci. 110:2573-2578(1997).
RN   [3]
RP   SEQUENCE OF 514-790 FROM N.A. (CENTROSOMIN A).
RX   MEDLINE=91277032; PubMed=1829085;
RA   Joswig G., Petzelt C., Werner D.;
RT   "Murine cDNAs coding for the centrosomal antigen centrosomin A.";
RL   J. Cell Sci. 98:37-43(1991).
RN   [4]
RP   REVISIONS.
RA   Joswig G., Petzelt C., Werner D.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC       METHIONYL-TRNAI AND MRNA.
CC   -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14172; AAA90910.1; -.
DR   EMBL; X84651; CAA59144.1; -.
DR   EMBL; X17373; CAA35246.1; -.
DR   PIR; S13800; S13800.
DR   MGD; MGI:95301; EIF3.
KW   Initiation factor; Protein biosynthesis; Repeat; Phosphorylation.
FT   DOMAIN      924   1133       21 X 10 AA TANDEM REPEAT OF D-[DE]-D-R-
FT                                [GP]-[PS]-[RW]-R-[GN]-[AM].
FT   CONFLICT    613    647       RQEAKEREKERILQEHEQIKKKTVRERLEQIKKTE ->
FT                                PRGKGAREGTNPSRTRANQEENCSGAVRADQEDR (IN
FT                                REF. 2 AND 3).
FT   CONFLICT    683    684       EL -> DY (IN REF. 2 AND 3).
FT   CONFLICT    717    717       Q -> H (IN REF. 2 AND 3).
FT   CONFLICT    766    766       A -> V (IN REF. 2 AND 3).
FT   CONFLICT    787    790       RHSR -> SIVA (IN REF. 3).
FT   CONFLICT    793    793       E -> D (IN REF. 2).
SQ   SEQUENCE   1344 AA;  161949 MW;  F4CAE2169F577712 CRC64;


  Query Match             19.9%;  Score 70.5;  DB 1;  Length 1344;
  Best Local Similarity   26.4%;  Pred. No. 5.2;
  Matches   19;  Conservative   20;  Mismatches   24;  Indels    9;  Gaps    3;

Qy       1 KRDPQQREYEDCRRRCEQQE-PRQQHQCQLRCREQQRQHGR--GGDMMNPQ------RGG 51
             ||: : |||::  :: |: |  ::| : ::  ||::|:  |  | | :: :      |
Db     836 KREEELREYQERVKKLEEVERKKRQRELEIEERERRREEERRLGDDPLSRKDSRWGDRDS 895

Qy      52 SGRYEEGEEEQS 63
             | : :| |  |
Db     896 EGTWRKGPEADS 907


RESULT  12
WT1_SMIMA
ID   WT1_SMIMA      STANDARD;      PRT;   239 AA.
AC   P49953;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   WILMS' TUMOR PROTEIN (FRAGMENT).
GN   WT1.
OS   Sminthopsis macroura (Stripe-faced dunnart).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sminthopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=96068905; PubMed=7478606;
RA   Kent J., Coriat A.M., Sharpe P.T., Hastie N., van Heyningen V.;
RT   "The evolution of WT1 sequence and expression pattern in the
RT   vertebrates.";
RL   Oncogene 11:1781-1792(1995).
CC   -!- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC       AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3' (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85732; CAA59737.1; -.
DR   HSSP; P08046; 1A1L.
DR   INTERPRO; IPR000822; -.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW   Transcription regulation; Alternative splicing; Anti-oncogene.
FT   NON_TER       1      1
FT   DOMAIN      113    228       ZINC FINGERS.
FT   ZN_FING     113    137       C2H2-TYPE.
FT   ZN_FING     143    167       C2H2-TYPE.
FT   ZN_FING     173    195       C2H2-TYPE.
FT   ZN_FING     204    228       C2H2-TYPE.
FT   VARSPLIC    198    200       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   239 AA;  27793 MW;  6707678A7259A624 CRC64;


  Query Match             19.4%;  Score 69;  DB 1;  Length 239;
  Best Local Similarity   25.0%;  Pred. No. 1.5;
  Matches   15;  Conservative   14;  Mismatches   25;  Indels    6;  Gaps    1;

Qy       4 PQQREYEDCRRRCEQQEPRQQHQCQ------LRCREQQRQHGRGGDMMNPQRGGSGRYEE 57
             | | :::|| ||  : :  ::|| :       :|:  ||:  |   :    | :|:  |
Db     142 PYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQCKTCQRKFSRSDHLKTHTRTHTGKTSE 201


RESULT  13
WT1_ALLMI
ID   WT1_ALLMI      STANDARD;      PRT;   288 AA.
AC   P50902;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   WILMS' TUMOR PROTEIN (FRAGMENT).
GN   WT1.
```

```
OS   Alligator mississippiensis (American alligator).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Crocodylidae; Alligatorinae; Alligator.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96068905; PubMed=7478606;
RA   Kent J., Coriat A.M., Sharpe P.T., Hastie N., van Heyningen V.;
RT   "The evolution of WT1 sequence and expression pattern in the
RT   vertebrates.";
RL   Oncogene 11:1781-1792(1995).
CC   -!- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC       AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85730; CAA59735.1; -.
DR   HSSP; P08046; 1A1L.
DR   INTERPRO; IPR000822; -.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW   Transcription regulation; Alternative splicing.
FT   NON_TER       1      1
FT   ZN_FING     162    186       C2H2-TYPE.
FT   ZN_FING     192    216       C2H2-TYPE.
FT   ZN_FING     222    244       C2H2-TYPE.
FT   ZN_FING     253    277       C2H2-TYPE.
FT   VARSPLIC    245    247       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   288 AA;  33111 MW;  33E26F7DBE7BFDDE CRC64;


  Query Match             19.4%;  Score 69;  DB 1;  Length 288;
  Best Local Similarity   25.0%;  Pred. No. 1.8;
  Matches   15;  Conservative   14;  Mismatches   25;  Indels    6;  Gaps    1;

Qy       4 PQQREYEDCRRRCEQQEPRQQHQCQ------LRCREQQRQHGRGGDMMNPQRGGSGRYEE 57
           | | :::|| ||  : :  ::|| :       :|:  ||:  |   :    |  :|:  |
Db     191 PYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQCKTCQRKFSRSDHLKTHTRTHTGKTSE 250


RESULT  14
WT1_RAT
ID   WT1_RAT        STANDARD;      PRT;   448 AA.
AC   P49952;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   WILMS' TUMOR PROTEIN HOMOLOG.
GN   WT1 OR WT-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RX   MEDLINE=93046155; PubMed=1330293;
RA   Sharma P.M., Yang X., Bowman M., Roberts V., Sukumar S.;
RT   "Molecular cloning of rat Wilms' tumor complementary DNA and a study
RT   of messenger RNA expression in the urogenital system and the brain.";
RL   Cancer Res. 52:6407-6412(1992).
CC   -!- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC       AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS.
CC   -!- TISSUE SPECIFICITY: KIDNEY (BY SIMILARITY).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.
CC   -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69716; CAA49373.1; -.
DR   TRANSFAC; T02352; -.
DR   INTERPRO; IPR000822; -.
DR   INTERPRO; IPR000976; -.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   PRINTS; PR00048; ZINCFINGER.
DR   PRINTS; PR00049; WILMSTUMOUR.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW   Transcription regulation; Alternative splicing; Anti-oncogene.
FT   DOMAIN       27     82       PRO-RICH.
FT   DOMAIN      322    437       ZINC FINGERS.
FT   ZN_FING     322    346       C2H2-TYPE.
FT   ZN_FING     352    376       C2H2-TYPE.
FT   ZN_FING     382    404       C2H2-TYPE.
FT   ZN_FING     413    437       C2H2-TYPE.
FT   VARSPLIC    249    265       MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT   VARSPLIC    407    409       MISSING (IN ISOFORM 2 AND ISOFORM 4).
SQ   SEQUENCE   448 AA;  49193 MW;  329AC9AC1FF73F76 CRC64;


  Query Match             19.4%;  Score 69;  DB 1;  Length 448;
  Best Local Similarity   25.0%;  Pred. No. 2.7;
  Matches   15;  Conservative   14;  Mismatches   25;  Indels    6;  Gaps    1;

Qy       4 PQQREYEDCRRRCEQQEPRQQHQCQ------LRCREQQRQHGRGGDMMNPQRGGSGRYEE 57
           | | :::|| ||  : :  ::|| :       :|:  ||:  |   :    |  :|:  |
Db     351 PYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQCKTCQRKFSRSDHLKTHTRTHTGKTSE 410


RESULT  15
WT1_HUMAN
ID   WT1_HUMAN      STANDARD;      PRT;   449 AA.
AC   P19544;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   WILMS' TUMOR PROTEIN (WT33).
GN   WT1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL KIDNEY;
RX   MEDLINE=90158822; PubMed=2154702;
RA   Gessler M., Poustka A., Cavenee W., Neve R.L., Orkin S.H.,
RA   Bruns G.A.P.;
RT   "Homozygous deletion in Wilms tumours of a zinc-finger gene
RT   identified by chromosome jumping.";
RL   Nature 343:774-778(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=92052142; PubMed=1658787;
RA   Haber D.A., Sohn R.L., Buckler A.J., Pelletier J., Call K.M.,
RA   Housman D.E.;
RT   "Alternative splicing and genomic structure of the Wilms tumor gene
```

```
RT   WT1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9618-9622(1991).
RN   [3]
RP   SEQUENCE OF 85-449 FROM N.A.
RX   MEDLINE=90150277; PubMed=2154335;
RA   Call K.M., Glaser T., Ito C.Y., Buckler A.J., Pelletier J.,
RA   Haber D.A., Rose E.A., Kral A., Yeger H., Lewis W.H., Jones C.,
RA   Housman D.E.;
RT   "Isolation and characterization of a zinc finger polypeptide gene at
RT   the human chromosome 11 Wilms' tumor locus.";
RL   Cell 60:509-520(1990).
RN   [4]
RP   IDENTIFICATION OF START CODON AND ALTERNATIVE SPLICE SITES.
RX   MEDLINE=91141522; PubMed=1671709;
RA   Buckler A.J., Pelletier J., Haber D.A., Glaser T., Housman D.E.;
RT   "Isolation, characterization, and expression of the murine Wilms'
RT   tumor gene (WT1) during kidney development.";
RL   Mol. Cell. Biol. 11:1707-1712(1991).
RN   [5]
RP   VARIANT WT CYS-366.
RX   MEDLINE=92279213; PubMed=1317572;
RA   Little M.H., Prosser J., Condie A., Smith P.J., van Heyningen V.,
RA   Hastie N.D.;
RT   "Zinc finger point mutations within the WT1 gene in Wilms tumor
RT   patients.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:4791-4795(1992).
RN   [6]
RP   VARIANTS DDS.
RX   MEDLINE=92005721; PubMed=1655284;
RA   Pelletier J., Bruening W., Kashtan C.E., Mauer S.M., Manivel J.C.,
RA   Striegel J.E., Houghton D.C., Junien C., Habib R., Fouser L.,
RA   Fine R.N., Silverman B.L., Haber D.A., Housman D.;
RT   "Germline mutations in the Wilms' tumor suppressor gene are
RT   associated with abnormal urogenital development in Denys-Drash
RT   syndrome.";
RL   Cell 67:437-447(1991).
RN   [7]
RP   VARIANTS DDS.
RX   MEDLINE=93265053; PubMed=1338906;
RA   Baird P.N., Santos A., Groves N., Jadresic L., Cowell J.K.;
RT   "Constitutional mutations in the WT1 gene in patients with
RT   Denys-Drash syndrome.";
RL   Hum. Mol. Genet. 1:301-305(1992).
RN   [8]
RP   VARIANTS DDS.
RX   MEDLINE=93271983; PubMed=8388765;
RA   Little M.H., Williamson K.A., Mannens M., Kelsey A., Gosden C.,
RA   Hastie N., van Heyningen V.;
RT   "Evidence that WT1 mutations in Denys-Drash syndrome patients may act
RT   in a dominant-negative fashion.";
RL   Hum. Mol. Genet. 2:259-264(1993).
RN   [9]
RP   VARIANT MESOTHELIOMA GLY-273.
RX   MEDLINE=94004972; PubMed=8401592;
RA   Park S., Schalling M., Bernard A., Maheswaran S., Shipley G.C.,
RA   Roberts D., Fletcher J., Shipman R., Rheinwald J., Demetri G.,
RA   Griffin J., Minden M., Housman D.E., Haber D.A.;
RT   "The Wilms tumour gene WT1 is expressed in murine mesoderm-derived
RT   tissues and mutated in a human mesothelioma.";
RL   Nat. Genet. 4:415-420(1993).
RN   [10]
RP   VARIANTS WT SER-181 AND ALA-253.
RX   MEDLINE=97268681; PubMed=9108089;
RA   Schumacher V., Schneider S., Figge A., Wildhardt G., Harms D.,
RA   Schmidt D., Weirich A., Ludwig R., Royer-Pokora B.;
RT   "Correlation of germ-line mutations and two-hit inactivation of the
RT   WT1 gene with Wilms tumors of stromal-predominant histology.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:3972-3977(1997).
RN   [11]
RP   VARIANTS DMS TYR-377; LEU-383 AND ASN-396.
RX   MEDLINE=98198341; PubMed=9529364;
RA   Jeanpierre C., Denamur E., Henry I., Cabanis M.-O., Luce S.,
RA   Cecille A., Elion J., Peuchmaur M., Loirat C., Niaudet P.,
RA   Gubler M.-C., Junien C.;
RT   "Identification of constitutional WT1 mutations, in patients with
RT   isolated diffuse mesangial sclerosis, and analysis of
RT   genotype/phenotype correlations by use of a computerized mutation
RT   database.";
RL   Am. J. Hum. Genet. 62:824-833(1998).
RN   [12]
RP   REVIEW.
RX   MEDLINE=92207913; PubMed=1313285;
RA   Haber D.A., Buckler A.J.;
RT   "WT1: a novel tumor suppressor gene inactivated in Wilms' tumor.";
RL   New Biol. 4:97-106(1992).
RN   [13]
RP   REVIEW.
RX   MEDLINE=93345769; PubMed=8393820;
RA   Rauscher F.J. III;
RT   "The WT1 Wilms tumor gene product: a developmentally regulated
RT   transcription factor in the kidney that functions as a tumor
RT   suppressor.";
RL   FASEB J. 7:896-903(1993).
CC   -!- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC       AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE KIDNEY AND A SUBSET OF
CC       HEMATOPOIETIC CELLS.
CC   -!- DISEASE: WILM'S TUMOR (WT) IS AN EMBRYONAL MALIGNANCY OF THE
CC       KIDNEY THAT AFFECTS APPROXIMATELY 1 IN 10,000 INFANTS AND YOUNG
CC       CHILDREN. IT OCCURS BOTH IN SPORADIC AND HEREDITARY FORMS.
CC       INACTIVATION OF WT1 IS ONE OF THE CAUSES OF WILM'S TUMOR.
CC   -!- DISEASE: PATIENTS WITH DENYS-DRASH SYNDROME (DDS), WHICH IS
CC       CHARACTERIZED BY A TYPICAL NEPHROPATHY AND GENITAL ABNORMALITIES,
CC       HAVE DEFECTS IN THE WT1 GENE.
CC   -!- DISEASE: DEFECTS IN WT1 ARE ALSO A CAUSE OF DIFFUSE MESANGIAL
CC       SCLEROSIS (DMS), A FORM A DDS.
CC   -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   -!- DATABASE: NAME=WT1; NOTE=WT1 mutation database;
CC       WWW="http://www.umd.necker.fr:2003/".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X51630; CAA35956.1; ALT_INIT.
DR   EMBL; M80232; AAA61299.1; -.
DR   EMBL; M80217; AAA61299.1; JOINED.
DR   EMBL; M80218; AAA61299.1; JOINED.
DR   EMBL; M80219; AAA61299.1; JOINED.
DR   EMBL; M80220; AAA61299.1; JOINED.
DR   EMBL; M80221; AAA61299.1; JOINED.
DR   EMBL; M80228; AAA61299.1; JOINED.
DR   EMBL; M80229; AAA61299.1; JOINED.
DR   EMBL; M80231; AAA61299.1; JOINED.
DR   EMBL; M30393; AAA36810.1; -.
DR   PIR; A34673; A34673.
DR   PIR; S08273; S08273.
DR   TRANSFAC; T00899; -.
DR   MIM; 194070; -.
DR   MIM; 194080; -.
DR   MIM; 256370; -.
DR   INTERPRO; IPR000822; -.
DR   INTERPRO; IPR000976; -.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   PRINTS; PR00048; ZINCFINGER.
DR   PRINTS; PR00049; WILMSTUMOUR.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW   Transcription regulation; Alternative splicing; Anti-oncogene;
```

```
KW   Disease mutation.
FT   DOMAIN       27     83       PRO-RICH.
FT   DOMAIN      323    438       ZINC FINGERS.
FT   ZN_FING     323    347       C2H2-TYPE.
FT   ZN_FING     353    377       C2H2-TYPE.
FT   ZN_FING     383    405       C2H2-TYPE.
FT   ZN_FING     414    438       C2H2-TYPE.
FT   VARSPLIC    250    266       MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT   VARSPLIC    408    410       MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT   VARIANT     181    181       P -> S (IN WT).
FT                                /FTId=VAR_007739.
FT   VARIANT     223    223       S -> N (IN WT).
FT                                /FTId=VAR_007740.
FT   VARIANT     253    253       G -> A (IN WT).
FT                                /FTId=VAR_007741.
FT   VARIANT     273    273       S -> G (IN MESOTHELIOMA).
FT                                /FTId=VAR_007742.
FT   VARIANT     330    330       C -> Y (IN DDS).
FT                                /FTId=VAR_007743.
FT   VARIANT     360    360       C -> G (IN DDS).
FT                                /FTId=VAR_007744.
FT   VARIANT     366    366       R -> C (IN WT).
FT                                /FTId=VAR_007745.
FT   VARIANT     366    366       R -> H (IN DDS).
FT                                /FTId=VAR_007746.
FT   VARIANT     373    373       H -> Q (IN DDS).
FT                                /FTId=VAR_007747.
FT   VARIANT     377    377       H -> Y (IN DDS/DMS).

  Query Match             19.4%;  Score 69;  DB 1;  Length 449;
  Best Local Similarity   25.0%;  Pred. No. 2.7;
  Matches   15;  Conservative   14;  Mismatches   25;  Indels    6;  Gaps    1;

Qy       4 PQQREYEDCRRRCEQQEPRQQHQCQ------LRCREQQRQHGRGGDMMNPQRGGSGRYEE 57
           | | :::|| ||  : :  ::|| :       :|:  ||:  |   :    |  :|:  |
Db     352 PYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQCKTCQRKFSRSDHLKTHTRTHTGKTSE 411


Search completed: March  1, 2001, 16:16:33
Job time: 413 sec
```